2020

Multimodal Vibrational Studies of Drug Uptake In Vitro: Is the Whole Greater than the Sum of Their Parts?

David Perezguaita
Technological University Dublin, david.perezguaita@tudublin.ie

Karolina Chrabaszcz
Jagiellonian University Cracow

Kamilla Malek
Jagiellonian University Cracow

See next page for additional authors

Recommended Citation

Perez-Guaita, D., Chrabaszcz, K., Malek, K.& Byrne, H.J. (2020). Multimodal vibrational studies of drug uptake in vitro: Is the whole greater than the sum of their parts? *Journal of BIOphotonics*, vol.13(12), e202000264. doi: https://doi.org/10.1002/jbio.202000264.

Funder: European Commission

MULTIMODAL VIBRATIONAL STUDIES OF DRUG UPTAKE IN VITRO: IS THE WHOLE GREATER THAN THE SUM OF THEIR PARTS?

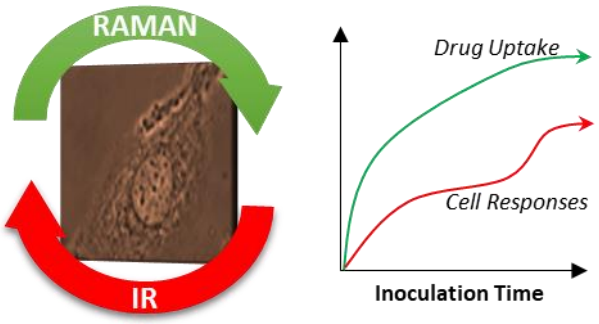
David Perez-Guaita*[1], Karolina Chrabaszcz[2], Kamilla Malek[2], Hugh J. Byrne[1]

[1]*FOCAS Research Institute, Technological University Dublin, City Campus, Dublin 8, Ireland*

[2] *Faculty of Chemistry, Jagiellonian University, Krakow, Poland*

ABSTRACT

Herein, we investigated the use of multimodal Raman and infrared (IR) spectroscopic microscopy for the elucidation of drug uptake and subsequent cellular responses. Firstly, we compared different methods for the analysis of the combined data. Secondly, we evaluated whether the combined analysis provided enough benefits to justify the fusion of the data. A459 cells inoculated with doxorubicin (DOX) at different times were fixed and analysed using each technique. Raman spectroscopy provided high sensitivity to DOX and enabled an accurate estimation of the drug uptake at each time point, whereas IR provided a better insight into the resultant changes in the biochemical composition of the cell. In terms of benefits of data fusion, 2D correlation analysis allowed the study of the relationship between IR and Raman variables, whereas the joint analysis of IR and Raman enabled the correlation of the different variables to be monitored over time. In summary, the complementary nature of IR and Raman makes the combination of these vibrational techniques an appealing tool to follow drug kinetics and cellular response.



Keywords: **Infrared, Raman, Multimodal Analysis, Drug uptake, DOX, Chemometrics**

1. Introduction

A clear understanding of drug modes and mechanism of action is essential for drug development and determination of clinical safety on both cellular and molecular levels. For that matter, it is important to recognise the pattern of biochemical interactions which occurs in a cell under a pharmacological effect, from the assimilation of the drug by the cell to subsequent changes in the metabolic pathways and phenotypes produced. [1,2]

The increase of restrictions in regulatory requirements in the EU and US (EU Directive-2010/63/EU and US Public Law 106-545, 2010, 106th Congress) has promoted the use of high throughput in-vitro investigations of modes and mechanisms of action. So far, the number of methodologies for identification of targets of drugs and their modes of action is still limited, however.[3,4] Omics technologies provide a wide range of approaches including lipidomic, proteomic and metabolomic techniques which enable the detection of several cellular biomolecules and a comprehensive analysis of the metabolic pathways altered by the drug. [5] However, in general they are time-consuming, need complex sample preparation and may destroy the sample. This makes it difficult to adapt these techniques as high-throughput methods to screen the mode and mechanism of action of the thousands of drug candidates developed yearly.

Raman and Infrared (IR) spectroscopies are rapid, label-free and non-destructive methods which provide molecular information about a biochemical composition of a sample. These techniques, when combined with imaging microscopic systems, are used to study the distribution of chemical substances in cells and tissues.[6,7] Raman microspectrocopy has been demonstrated to be capable of monitoring sub-cellular drug accumulation and distribution of doxorubicin (DOX) in lung cancer cell lines and to identify the mechanism of the DOX action and cellular resistance.[8][9] With a spatial resolution similar to that of confocal laser scanning fluorescence microscopy [10,11], confocal Raman microspectroscopy indicated an initial accumulation and saturation of the drug in the nucleoli (~6 hrs), followed by nucleus of the cells (~12 hrs), before it began to accumulate in the cytoplasm (>24 hrs)[10,11] . 2D Correlation[12], as well as Multivariate Curve Resolution-Alternating Least Squares analysis [9] indicated a simultaneous evolution of spectral features associated with RNA in the nucleolus and DNA in the nucleus of the cells, associating the spectral signatures with the chemical binding, and therefore mechanism of action of the drug. Employing the median cytotoxic dose (IC_{50}), as determined using the 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) cytotoxicity assay, differences in the long term cellular responses [11] of different cell lines were associated with cell recovery and anti-apoptotic mechanisms [13].

On the other hand, IR has been successfully employed to follow molecular changes induced by chemotherapy drugs.[14] Similar to Raman microspectroscopy, the combination of IR spectroscopy and principal component analysis

(PCA) [15] or partial least squares (PLS) [16] has been proposed to distinguish the mode of action displayed by chemotherapeutics such as tamoxifen, toremifene, imatinib, and DOX.

Both, IR and Raman spectra provide information about the vibrational modes of molecules. However, they differ significantly in the mechanism of excitation from the ground state to the excited state and thus molecules can show very different activity for Raman and IR spectroscopy.[17] In IR, polar bonds from asymmetric molecules show strong bands, whereas symmetric and highly conjugated molecules are very active in Raman. In addition, Raman associated phenomena such as resonance or pre-resonance makes this technique particularly sensitive to molecular groups with electronic transitions close to the laser wavelength, generally in the visible and near IR range (e.g. chromophores).

These differences make the techniques complementary in the investigation of complex matrices such as biological samples. For example, haemoglobin within red blood cells has been widely studied by both techniques. While the IR spectra is clearly dominated by the amide bands of the protein backbone,[18] the Raman spectrum is completely different, showing bands related to the heme group.[19] In this regard, cells and tissues are composed of a wide variety of Raman and IR active molecules and the complementarity of the techniques can be exploited to obtain a more comprehensive analysis. For example, some (conjugated) drugs are very strong Raman scatterers and therefore easily tracked by Raman. DNA/RNA conjugated bases also show relatively strong Raman signals, but proteins show relatively stronger in IR. Furthermore, subcellular resolution is more readily accessible with Raman microscopy, whereas IR is more readily applicable to cell population studies. [20]

Indeed, there is not a single technique or analytical configuration which can provide information of all the biomolecules from a clinical sample. The combination of different techniques which cover specific parts of the phenotype (e.g. proteome, lipidome and metabolome) is gaining popularity, allowing capture of a wider range of biomarkers and thus a more representative picture of the molecules involved.[21] The price to pay is the considerably increased complexity of the data analysis. The multimodal study of a biological process implies the simultaneous analysis of responses from different origins, which demands new ways to analyse data considering the relationship between the information and biomarkers addressed by the different modalities. In recent years, data fusion methods have been developed to study correlations between variables obtained using different spectroscopic methods,[22] as well as to draw comprehensive conclusions from extended datasets.[23]

In the case of vibrational techniques, IR or Raman spectra are already intrinsically complex, being composed by multiple overlapping bands representing the concentration of the multitude of complex molecules within a sample. Hence, the combination of Raman and IR variables becomes a challenge. Different methods have been applied to

integrate data from IR, Raman and other metabolic techniques, including PCA [24,25], 2D correlation analysis [26,27], MCR-ALS [28] and joint individual variation analysis [29]. In the context of a drug uptake study, the analysis becomes even more complex when the incubation time variable is introduced in the system. In order to describe the drug kinetics as well as the responses of the cell expressed in the IR and Raman spectra, the correlations among the different variables should be studied across the duration of drug uptake by, and response of the cell.

The aim of this academic exercise was to evaluate the simultaneous use of IR and Raman spectroscopy to study drug uptake and cellular response. To achieve this aim, we probed human lung carcinoma cells inoculated with DOX at different timepoints by means of near infrared Raman and High Definition (HD) Fourier Transform (FT)IR microspectroscopy. The study addressed (i) which methods are the most suitable for the analysis of the combined data and (ii) whether the joint analysis provided extra information in comparison with the individual analysis of both techniques *and (iii) the usefulness of incorporating IR data considering that Raman is a well-established technique for the analysis.*

2. Methods

2.1 Cell Culture

Adenocarcinomic human alveolar basal epithelial A549 cells were cultured as in [11], using DMEM-F12 with 10% fetal bovine serum (FBS) at 37 °C in a humidified atmosphere containing 5% CO_2. "Cells were subcultured regularly (2-3 days) in order to maintain a confluency of ~60% ".. Some aliquots of cells were seeded on CaF_2 windows (~1 × 10^4 /window) from Crystan Ltd (Dorset, UK) and incubated for 24 h. The medium was then removed, and samples were rinsed three times with PBS. Then the cells were incubated with DOX at the median inhibitory concentration (IC_{50}), adjusted to the cell number, for different time periods (0.5, 1, 2, 4, 7, 10, 20 and 48 h) [11] incubation. The IC_{50} cytotoxicity concentration used (0.55 µM) was previously determined using a 3-[4,5-dimethylthiazol-2-yl]-2,5-diphenyl tetrazolium bromide (MTT) assay (24 hours of incubation time).[11] The cells were then washed three times with sterile PBS before they were fixed in formalin (4%, 15 min). Prior to the analysis, samples were rinsed with ultrapure water.

2.2 FT IR spectroscopic imaging

Cells deposited on CaF_2 windows were measured by FTIR microscopy using a Agilent 670-IR FTIR spectrometer and a 620-IR microscope (Santa Clara, California) equipped with focal plane array (FPA) detector cooled with liquid nitrogen. The detector consists of matrix of 16 384 pixels, arranged in a 128 × 128 grid format. IR images were collected in transmission mode. High Definition (HD) FTIR spectroscopic imaging used a ×15 Cassegrain objective and

condenser optics with NA of 0.62 and a projected FPA pixel size of 1.1 μm × 1.1 μm giving a measured area of 141 μm × 141 μm. A high magnification accessory was use giving a total IR system magnification of 36× with a spatial resolution of 3.0 μm at 2500 cm^{-1}. All FTIR spectra were recorded by co-adding of 128 scans and in the range of 3800 to 900 cm^{-1} with a spectral resolution of 4 cm^{-1}. For each sample two images were obtained, covering 3-4 cells each. The pixels within a cell, covering both cytoplasm and nucleus, were manually selected and averaged to create a dataset containing 5-8 spectra per cell for each timepoint.

2.3 Raman

Raman analysis was performed after the FTIR analysis in order to avoid any problems related to laser damage. Raman spectra were obtained using a Jobin Yvon LabRAM HR 800 spectrometer from Horiba (Kagawa, Japan) equipped with a Peltier cooled 16-bit CCD, 300 lines/mm grating and 100 μm confocal hole and a 785 nm, 300 mW diode laser as source (100% of laser power), *resulting in ~70mW on the sample*[11,13]. Spectra were acquired directly in dry conditions, using a ×100 objective (LCPlanN, Olympus) in the 400 cm^{-1}- 1800 cm^{-1} range. For each cell, the average of 3 spectra obtained during 30s was computed. Spectra were randomly obtained from the cell region covering both nuclei and nucleoli. The final data set comprised spectra of 20-30 cells for each time point.

2.4 Data Analysis

Data preprocessing and analysis was performed using Matlab from Mathworks (NAtik, USA). Spectral pre-processing, PCA and orthogonal partial least squares regression (oPLSR) were performed employing functions of the PLS-Toolbox library from Eigenvector Research Inc (Mason, USA). The synchronous spectrum was calculated using the function *fft2dcorrelation* from the MIDAS software[30]. MCR-ALS was obtained using the MCR-ALS GUI toolbox.[31] The scripts detailing the data analysis as well as the data are available in the Zenodo repository (10.5281/zenodo.3784974). Raman spectra were baselined using the "Weighted least squares" method from PLS-toolbox (3-degree polynomial), normalised using the amide I band (1600-1700 cm^{-1}) and smoothed using the Savitzky-Golay algorithm, using a 9 points window. IR spectra were baselined using a first order degree polynomial fitted on the 1790-1780 and 950- 920 cm^{-1} regions, smoothed (Savitzky-Golay algorithm, 7 points window) and normalised with the "Standard Normal Variate" method.

3. Results

3.1 Monitoring drug uptake using Raman spectroscopy

Figure 1a depicts the average (n=30) Raman spectra from cells fixed at different incubation times. The spectra show typical bands from cells, including protein bands corresponding to Amide I (~1655 cm^{-1}) and Amide III vibrations (~1242 cm^{-1}) [32], and lipid bands, such as the C-H deformation found at ~1441 cm^{-1}[33], In addition, the band found at ~785 cm^{-1} can be assigned to cytosine and thymine, while that found at 816 cm^{-1} is assigned to the O–P–O stretching of RNA [8]. As the incubation time increases, several additional bands start to emerge. These correspond to vibrational modes of DOX, such as the C-C-O (~436 cm^{-1}) and C-O (~457 cm^{-1}) vibrations,[34] which are clearly spectrally isolated from the cell bands. Figure 1c shows the spectrum of a standard of pure DOX in aqueous solution, which indicates the presence of additional DOX bands, which overlap the cell features, notably the bands found at ~1212 and ~1245 cm^{-1}, related to C-O-H and C-O groups, respectively.[8]

The integrated value of the bands found in the range 430-482 cm^{-1} was used to follow the uptake of DOX by the cell as a function of incubation time (See Figure 1b). Values initially increased sharply, before reaching a plateau at ~ 7-10 hours, indicating saturation of DOX inside the nucleus. This observation is in good agreement with previous studies[8], which indicated that the accumulation in the nucleolus and nucleus of the cells began to saturate at times of ~6 and ~12 hours, respectively.". It has to be noted that, in the current study, measurements over both nucleolus and nucleolus were integrated, which could explain the fact that the saturation time is in the middle of the interval previously reported for each subnuclear region measured independently. Furthermore, the large standard deviation found in this study was presumably caused by the lack of synchronization of the cells.

PCA was performed to investigate multivariant trends in the data. The score values of the first two Principal Components, PC1 and PC2, showed spectral variations which were not correlated with time (See SM1). In contrast, the scores of the third PC showed a similar trend to that of the integrated DOX band (See Figure 1d). Moreover, the spectral loading of PC3 (See Figure 1c) shows identical features to the DOX standard, indicating that PC3 captured, to some extent, the variance associated with the introduction of the drug. Other minor peaks can be also observed, including the negative band found at 815 cm^{-1}, which could indicate of the intercalation of the drug and nucleic acids [8,35]. DNA bands Negative bands in the 1250-1400 cm^{-1} region can be also observed.

oPLSR was employed to regress the Raman spectra against incubation time, in order to acquire information about the contribution of each band to the kinetic process, but showed some limitations related to the fact that the relationship between the time and the DOX concentration was not linear. This can be observed in the poor correlation between the cross-validated (CV) predicted and actual time values (See Figure 1f) plot. This behaviour has been previously discussed with simulated data[9] and evidences the problems of using a linear model in a time dependent data unless the kinetic

process is separated in linear trams. Nevertheless, the Root mean square error of cross validation (RMSECV) error obtained (10.39 h) using 4 latent variables (LV) was deemed significant by a permutation test ($p<0.005$)[36], showing some correlation between the incubation time and the Raman spectra. Considering this value and that the correlation was observed in an unsupervised manner, spectral makers were investigated in the oPLSR model. The importance and direction of each variable in the model was studied using the selectivity ratio, a measure of the usefulness of each variable, and the component loading of the oPLSR[37], respectively. Figure 1e shows the average Raman spectra as a scatter plot, in which the size of each variable is proportional to the selectivity ratio and the colour corresponds to the loading of the first LV, which accounts for the variation correlated with the variable to predict. The bands found at 436, 461 and 1212 cm^{-1} showed both a high selectivity ratio and a high and positive loading value, indicating that the model used these bands to predict the time evolution. Furthermore, bands in the PC 1 loading showed similar shape than the DOX standard, suggesting that the intercalation of DOX with DNA did not perturb considerably the chemical environment of the drug.

Raman spectroscopy is a well stablished technique for the evaluation of the drug uptake and cellular responses. In this study, Raman data showed similar patterns to those found in previous studies, demonstrating that multivariate analysis of Raman data provides useful information about the kinetics of DOX uptake by and initial interactions within the cells. The following section describes an analogous analysis using IR data in order to investigate changes in the chemical composition associated with the drug uptake.

3.2 Evolution of the IR spectrum with the incubation time.

Figure 2a depicts the average IR spectra of whole cells fixed after different incubation times, in which typical features of biological samples can be seen. Spectra are dominated by protein amide II (~1540 cm^{-1}) and I (~1650 cm^{-1}) bands [38,39], as well as phosphate stretching bands from nucleic acids (~1241 and ~1074 cm^{-1})[40]. The amide I could also contain contributions from the set of broad DNA and RNA bands between 1 ~1600 and 1720 cm^{-1}[40], the latter being more clearly visible in Figure 2b. Lipid bands corresponding to the stretching mode of the ester C=O group and the bending mode of CH_2 can be found at ~1740 and ~1455 cm^{-1},[41,42] respectively. In addition, the stretching band of the COO^- group of fatty acids and amino acid side chains can be observed at ~1400 cm^{-1}. [41] Furthermore, the 3050-2800 cm^{-1} region presents bands associated with the stretching vibrations bands of C-H, corresponding to proteins, carbohydrates and lipids. [41] In particular, sharp bands found at ~2850 and ~2929 cm^{-1} correspond to the symmetric and asymmetric stretching of CH_2 and are in general associated with the presence of alkyl chains of lipids.

In comparison with the Raman analysis, no change of spectral profile as a function of incubation time is immediately obvious, on visual inspection. Similarly, PCA did not show any PC associated with the drug uptake, all the score values of the PCs being poorly correlated in any linear or non-linear fashion with the incubation time (See First 3PCs in SM2). This observation indicated that, either the incubation time does not alter the spectra (i.e. changes on the biochemical composition are below the limit of detection of IR) or the variability associated is not a main source of spectral variation and was captured by a combination of different PCs. In theory, supervised methods such as oPLSR could deal with the latter, the main drawbacks being potential problems of overfitting and other pitfalls, which could lead to overoptimistic results if internal validation (e.g. CV) is used. Results of the oPLSR modelling using 4 LV of the incubation time are depicted in Figure 2b and 2c. The actual versus CV-predicted incubation time plot showed a similarly poor correlation observed for the Raman data, with a RMSECV value of 8.1 h. Although the permutation test computed acceptable significance values ($p < 0.02$) for the RMSECV, considering that oPLSR was used to model a non-linear relationship, as well as the lack of any evidence of spectral changes associated with the incubation time found by the PCA, it is unrealistic to extract biochemical information from the model.

As a highly polarisable, conjugated molecule, DOX is a strong Raman scatterer, but in comparison, a relatively weak IR absorber, and thus it is not unexpected that its spectral features are more strongly evident in the Raman spectral profile of the cell, whereas it is "invisible" in the IR spectrum. Nevertheless, the Raman signals from the previous experiment could assist in extracting biochemical information from the IR spectrum. The integration values of the DOX bands in the 430-480 cm^{-1} range are a good indicator of the uptake of the drug by the nucleus, so this value can be used as the output variable in oPLSR model to correlate IR variables with the presence and action of the drug in the cell. The actual versus predicted plot (See Figure 2e) indicates a much better prediction performance compared with models for the prediction of the incubation time for both IR and Raman (Figure 1(f), Figure 2(c)). The permutation test (6LV) showed that the RMSECV found (0.025 h) was considerably more significant ($p<0.00001$) than the error calculated for the regression against incubation time. The fact that the IR variables could clearly predict the intensity of a DOX band in the Raman spectra indicated that there was a linear relationship between the infrared variables and the concentration of DOX in the nucleus of the cell. Two explanations are proposed: (i) The IR spectra contained drug bands and the concentration of DOX was above the limit of detection of technique. (ii) The drug provoked immediate changes in the biochemical constituents of the cells, which were directly proportional to the presence of DOX and which caused significant changes to the IR spectra.

Once again, the selectivity ratio and the component loading values were used to study the bands responsible for the prediction (See Figure 2d). Strong positive contributions (yellow colours) with high selectivity ratio values were found

at wavenumbers related to lipids such as 1400, 1740, 2850 and 2929 cm^{-1}. A strong positive contribution for the amide II (~1545 cm^{-1}) and a negative contribution for the amide I (~1650 cm^{-1}) were also observed, which may indicate changes in the conformation of proteins. The latter band could be also associated with a decrease of the RNA and DNA bands when increasing the amount of DOX in the nucleus. On the other hand, specific markers of DOX,[34] such as the bands found at 1283 cm^{-1} (δO-H …O) and 990cm^{-1} (δC-OH) were not evident. These bands are the most intense in the IR spectrum of DOX [29], and are in a region where other cell components do not show high absorbances. The lack of spectral markers associated to these bands evidences a negligible contribution of DOX bands to the IR spectra, probably because of the low concentration of DOX within the cell (i.e. below the limit of detection of IR). The absence of spectral markers associated with these bands evidences a negligible contribution of DOX bands to the IR spectra, probably because of the low concentration of DOX within the cell (i.e. below the limit of detection of IR). The concentration of a human genome within a cell is also relatively low, weighting ≈ 6pg[43] from the total weight of ≈1ng[44] of a human cell. In a previous study, Agudelo et al. identified changes in the DNA IR bands in transmission measurements of solutions of mixtures of DOX and DNA. [45] However, results of this study did not indicate that the interaction DNA-DOX affected DNA bands in the IR spectrum. This can be caused by the low intensity of DNA bands in the IR spectrum of cells, especially in fixed or dried cells, [40] when using such small pathlengths such as in the microscopy measurements of cells.

Hence, the analysis of variables supports the hypothesis that, rather than the drug, the IR spectroscopy detected changes produced by the presence of the drug, indicating that, as expected, the pre-resonant signal of DOX in Raman provides better sensitivity than IR in detecting the drug, which offers further information about changes in the biochemical composition.

3.3 2D Correlation Analysis.

Since the use of Raman data was beneficial in the analysis of IR variables in the oPLSR models, dedicated data fusion methods were investigated. 2D correlation analysis [46] was therefore employed to investigate the relationship between the IR and Raman variables. In brief, 2D correlation analysis computes a measure of the correlation for each combination of variables obtained from two different modalities, which are plotted in a correlation map to study the relationship between the information provided by the techniques across the different samples. 2D correlation provides a synchronous and an asynchronous spectrum. The first computes the correlation among the variables in the dataset, enabling the investigation of relationship of the IR and Raman intensities. The later takes into account the variable time, and can be

used to investigate the directions of the spectral changes across time. Further information about 2D correlation analysis can be found here.[26,27,46] To aid the interpretation of the synchronous coefficients, a simulated dataset was created, in a similar way to previous studies [9,12]. In this way, one can study the expected results and then compare them with the real data. The simulated Raman dataset from ref [9] was used, but it was extended to simulate also IR spectra. The Raman spectra contained two artificial features, which represent the initial rapid uptake and binding of the drug to RNA/DNA in the nucleus, and a slower, secondary response to represent the subsequent cellular response (See SM3). Similar constructs were also added to the IR spectra following the same kinetic profile. The drug uptake was represented as a decrease of typical DNA bands at 1660, 1239, 1085 cm^{-1} and, following the results of the oPLSR, the secondary cell response was simulated by an increase of lipid bands at 2930, 2850, 1750 and 1399 cm^{-1}. The 2D correlation analysis of the data obtained (See Figure 3a) showed two clear set of correlations between the Raman and IR bands which increased or decreased as a function of the incubation time. In particular, DOX Raman bands were positively correlated with the lipid IR bands (Red Arrows) and were inversely correlated with the DNA IR bands (Black Arrows).

Figure 3(b) shows the contour plot of the synchronous correlation of the IR and Raman spectral datasets measured at the different incubation times. It can be seen several negative and positive correlation. A large part of them correspond to correlations between DOX Raman bands and sets of IR bands bands, highlighting characteristic bands at 430-480 and ~1200 $cm^{-1,}$ which were found to be positively correlated with the IR bands related to lipids, found at ~2930, ~2850 and ~1400 cm^{-1} (See red arrows). The latter can be assigned to the COO^- stretching from fatty acids, but can also be assigned to amino acid side chains[41]. It has to be noted, however, that the characteristic ester band at 1740 cm^{-1} was not found to be positively correlated to the drug bands. Furthermore, in good agreement with the results from the oPLSR model, DOX bands were positively correlated with the band found at ~1540 cm^{-1} and negatively correlated with the band at ~1660 cm^{-1} (See black arrows). It can be also observed that the amide I band in both IR and Raman are positively correlated. Compared to the simulated data, the band at 1640 cm^{-1} showed the same negative correlation, but no correlation was found for the phosphate bands at 1239 and 1085 cm^{-1}. This supports the assignment of the 1660 cm^{-1} to changes in the protein conformation. On the other hand, the lipid bands at 2930, 2850 and 1400 cm^{-1} showed similar correlation and the band at 1750 cm^{-1} did not show any correlation. Once again, the Raman bands were helpful in identifying IR spectral markers linked to the presence of the drug.

The asynchronous correlation for the simulated and real dataset is shown in SM4. The analysis of the simulated data showed positive peaks between the Raman DOX bands and the FTIR lipidic bands. The real dataset showed a large number of positive and negative peaks. Considering the high degree of complexity of the system and the low number of

timepoints, the interpretation of such number of peaks in an asynchronous analysis is challenging and thus, only the synchronous correlation was analysed.

3.4 MCR-ALS

The main problem on studying correlations between variables is that they do not account for possible non-linear effects found in the kinetic processes of drug uptake. MCR-ALS is a more sophisticated method which enables analysis of complex variations of components within a chemical system. It is a versatile technique which can use constraints based on known information about the system to improve the interpretability of the results. In a previous study, this method was applied to simulated spectra of DOX, successfully capturing two different process with two different kinetics, which simulated the drug uptake and the subsequent cellular response[9]. In this study MCR-ALS, was applied to the combined IR and Raman dataset to decompose the extended data matrix of the concentration and spectra matrices, containing concatenated IR and Raman spectra. Based on a Singular Value Decomposition, three components were chosen. The uptake of the drug and its kinetics were incorporated into the model by constraining the first component to represent the Raman spectrum of the drug measured in aqueous solution, leaving the IR spectrum to be modelled by the MCR-ALS. Furthermore, the concentration profile of the first component was adjusted to the kinetic profile of DOX, measured as the integration of the Raman bands of DOX found at 430-480 cm^{-1}. In order to avoid possible propagation of experimental errors from the Raman to the IR, the concentration was fitted to a kinetic equation previously proposed.[12]. In summary, concentrations and Raman part of component #1 were constrained. The IR part of component #1, the IR and Raman spectra as well as the concentrations profiles of components #2 and #3 were also left without any equality constraint, to be modelled by the MCR-ALS. Furthermore, non-negativity constrains were applied for both concentrations and spectra.

Results of the MCR-ALS are depicted in Figure 4. The concentration profile of the three components (See Figure 4c) shows that component #1 (Blue) follows the kinetic pattern of the uptake of the drug, as described by the Raman spectra (introduced in the constraint). The other two components did not follow any defined trend over time. The Raman spectra of component #1 (See Figure 4a) also shows the constrained DOX standard spectrum, whereas the other components showed similar bands to the biological components of cells (Raman spectra of cell is indicated as a black dotted line). In the IR spectrum (See Figure 4b), component #2 and #3 also showed typical spectral features of cells, including amide and phosphate bands. It should be noted that these components show typical biochemical bands for both, Raman and IR, and considering the random fluctuations of their concentration profiles, they may represent changes of cell components not related to the time. The IR spectral features of component #1 are associated with the uptake and

represents the changes in the IR as a result of the action of DOX in the nucleus. They include bands found at 2929, 2850, 1740 and 1400 cm^{-1}, which can be related to lipids. In addition, a band at 1540 cm^{-1}, tentatively assigned to Amide II from proteins, can be found. The presence of this feature indicates an increase of this bands following the same kinetic as DOX uptake measured by the Raman signal. Compared to oPLSR and 2D correlation, similar lipidic features were found, although MCR-ALS also identified the 1740 cm^{-1} C=O band as associated with the uptake of DOX. The absence of this band in oPLSR and 2D correlation can be caused by negative contributions of the band at 1660 cm^{-1}, which could be associated to changes in protein conformation. In contrast, negative correlations of 1660 cm^{-1} were not identified by MCR-ALS, consistent with previous studies which evidenced some limitations in reporting synchronous negative and positive contributions. [9].

3.5 Evaluation of independent and joint analyis of IR and Raman data

In this study, we explored the analysis of DOX uptake by cells using IR and Raman spectroscopy, independently, and in a combined, or joint fashion. As stressed in the introduction, IR and Raman provide complementary information about the sample because, among other reasons, in most biomolecules there is an inverse correlation found between the Raman scattering cross section and the infrared extinction coefficient. Raman provided clear information about the drug uptake using typical bands at 430-480 cm-^{1}. The high sensitivity of Raman to detect DOX made the analysis straightforward, without the need of any supervised method to follow the DOX kinetics. However, the investigation of further changes to other biochemical composition was challenging. Raman is a well-established technique that may identify both drug uptake and subcellular response as it is also sensitive to changes in the biochemical composition of the cell. In this case, the subcellular response in Raman was limited to changes in the DNA bands. Regarding IR, the spectral contribution of DOX was negligible compared to that of the majority constituents such as proteins, carbohydrates, nucleic acids and lipids, which show characteristic spectral markers in the IR. The results showed that IR bands better illustrated changes in these majority components, evidencing changes in lipid and protein bands. Nevertheless, the investigation of the temporal evolution of such markers was troublesome because they were only found using supervised, linear methods such as oPLSR. The linear model was skewed, considering the fact that the drug uptake is not a linear function of incubation time. In the absence of previous analysis which confirmed the correlation of these bands with the incubation time (such was the case of Raman), the validity of these IR markers is questionable.

However, the use of the Raman signal of DOX as a variable to guide the oPLSR produced much more reliable results. It provided a response which was linear with the biochemical changes produced by the presence of the drug and enabled identification of IR spectral markers. This clearly exemplifies how the hyphenation of the techniques makes

facilitates the data-mining of information from one single technique by indirectly increasing the knowledge of the system. Furthermore, joint analysis of Raman and IR data using 2D correlation and MCR-ALS offered a much clear description of the system, showing strong correlations between DOX Raman bands and IR spectral markers associated with changes in biochemical constituents related to the presence of the drug. The joint analysis of the IR and Raman data was limited by the intrinsic constraints of the respective techniques. In the first place, the analysis was performed on fixed cells. The opacity of water in the mid-IR region makes the acquisition of spectra of living cells difficult. In Raman, this task is in theory more achievable, although special equipment is needed to maintain the cells in conditions similar to the ones found in an incubator. In addition, the lack of high throughput IR and Raman microscopes which can measure several samples containing thousands of cells automatically makes the acquisition of a representative number of cells for a large number of timepoints extremely time-consuming. This hampers the use of MCR-ALS to capture different kinetic process at different rates.[9] Finally, the differences in spatial resolution played a critical role in the interpretation of results. The IR spectra were averaged from the whole cell, reflecting changes in major components from nucleolus and cytoplasm. In contrast, Raman spectra were obtained from the nucleus and thus, the correlation between the DOX signals and IR markers of lipids should be understood as an increase of lipids in the whole cell as the concentration of DOX builds up in the nucleus. Recent technical advances in IR and Raman spectroscopy are aimed to overcome these challenges, for example increasing the spatial resolution of IR images by photothermal spectroscopy of live cells, simultaneously collected with Raman data [47], and development of rapid microfluidic cell-sorting systems to trap cells while maintaining suitable environmental conditions for cell survival[48]. The combined use of these new techniques with data fusion methodologies could provide a very valuable high content, high throughput tool for comprehensive analysis of, for example, drug uptake and cellular responses in pre-clinical drug development workflows.

As stressed in the introduction, the combination of both techniques increases considerably the complexity of the data to analyse and entails challenges in the measurement process (e.g. compatibility of substrates and spatial resolution). Thus, it has to be noted that the multimodal approach will not be the most suitable in all scenarios. In the case of this study, if the goal had been limited to the kinetic study of the drug uptake, localisation, and its interaction with the DNA, Raman alone would have provided a satisfactory analysis without the need for more complex analysis. However, if the goal is to obtain an extended picture, results demonstrated that the IR could provide information about the macromolecular changes associated with the drug uptake.

4. Conclusions

The study of DOX uptake and cellular response combining IR and Raman offered more valuable information than the independent analysis of data from the two techniques. Raman DOX bands found at 1212 and 430-480 cm^{-1} provided good insight into the drug uptake, due to the strong polarisability and conjugated bonds of DOX. Thus, Raman analysis provides kinetic information about the drug uptake as well as hints about DNA intercalation by itself without the need of complex mutivariate processing. In contrast, further information about the biochemical response of the cell was not easily elucidated. In the case of IR, spectral changes related to the incubation time were not found unless supervised analysis was employed.

In general, most of these band markers including the 2830, 2940, 1740 and 1400 cm^{-1} indicate an increasing on the relative concentration of lipids in cells correlated with an increase of the drug within the nucleus. The investigation of the changes in bands found at 1660 cm^{-1} (negatively correlated with the drug) and 1540 cm^{-1} (positively correlated with the drug) is more challenging. It can be tentatively assigned to changes in the ratio amide I to amide II provoked by changes in the conformation of proteins. Pleated sheets have been proposed as apoptosis marker in membrane of cells[49]. On the other hand, changes in the band found at 1660 cm^{-1} can also be caused by alterations in the conformation on the DNA, although typical phosphate DNA bands found at 1075 cm^{-1} and 1240 cm^{-1} did not show major changes. Thus, the interpretation of the bands found at 1540 and 1660 cm^{-1} remain inconclusive and further studies with a large number of timepoints should be carried out to study complex cell responses using MCR-ALS. In summary, the fusion of vibrational data improved the whole analysis in two main respects. Firstly, Raman assisted in extracting information from the IR spectra by providing knowledge about the drug uptake. Furthermore, the analysis of joint data by 2D correlation and MCR-ALS provided a comprehensive picture of the uptake process, combining the sensitivity of Raman for the drug detection and the information about the major biochemical components found in the IR bands. Although the multimodal vibrational analysis of drug uptake still suffers from limitations regarding the low spatial resolution of conventional IR microscopy as well as the general throughput of current commercial microscopes, the study demonstrates the potential of joint analysis approaches, in anticipation of emerging technological developments.

Acknowledgment

DPG acknowledges support from the European Research Council MSCA grant (Spectro-metrics, 020-MSCA-IF-2017 Project ID:796287). **K.M.** and **K.Ch.** acknowledge the financial support from the National Science Centre, Poland (UMO 2016/23/B/NZ4/01379).

[1] M. Schenone, V. Dančík, B. K. Wagner, P. A. Clemons, *Nat. Chem. Biol.* **2013**, *9*, 232.
[2] W.-L. Ng, K. M. Kazmierczak, G. T. Robertson, R. Gilmour, M. E. Winkler, *J. Bacteriol.* **2003**, *185*, 359.
[3] L. B. Tulloch, S. K. Menzies, R. P. Coron, M. D. Roberts, G. J. Florence, T. K. Smith, *IUBMB Life* **2018**, *70*, 9.
[4] T. Wecke, T. Mascher, *J. Antimicrob. Chemother.* **2011**, *66*, 2689.

[5] H. Matthews, J. Hanison, N. Nirmalan, *Proteomes* **2016**, *4*, 28.
[6] Z. Farhane, H. Nawaz, F. Bonnier, H. J. Byrne, *J. Biophotonics* **2018**, *11*, e201700258.
[7] K. Augustyniak, K. Chrabaszcz, A. Jasztal, M. Smeda, G. Quintas, J. Kuligowski, K. M. Marzec, K. Malek, *J. Biophotonics* **2019**, *12*, UNSP e201800345.
[8] Z. Farhane, F. Bonnier, O. Howe, A. Casey, H. J. Byrne, *J. Biophotonics* **2018**, *11*, e201700060.
[9] D. Perez-Guaita, G. Quintas, Z. Farhane, R. Tauler, H. J. Byrne, *Talanta* **2020**, *208*, 120386.
[10] Z. Farhane, F. Bonnier, A. Casey, H. J. Byrne, *The Analyst* **2015**, *140*, 4212.
[11] Z. Farhane, F. Bonnier, H. J. Byrne, *Anal. Bioanal. Chem.* **2017**, *409*, 1333.
[12] H. J. Byrne, F. Bonnier, Z. Farhane, *J. Biophotonics* **2019**, *12*, e201800328.
[13] Z. Farhane, F. Bonnier, M. A. Maher, J. Bryant, A. Casey, H. J. Byrne, *J. Biophotonics* **2017**, *10*, 151.
[14] P. L. Fale, A. Altharawi, K. L. A. Chan, *Biochim. Biophys. Acta BBA - Mol. Cell Res.* **2015**, *1853*, 2640.
[15] A. Altharawi, K. M. Rahman, K. L. A. Chan, *Analyst* **2019**, *144*, 2725.
[16] A. Derenne, R. Gasper, E. Goormaghtigh, *Analyst* **2011**, *136*, 1134.
[17] M. Diem, J. M. Chalmers, P. R. Griffiths, Eds., *Vibrational spectroscopy for medical diagnosis*, John Wiley & Sons, Chichester, England ; Hoboken, NJ, **2008**.
[18] D. Perez-Guaita, D. Andrew, P. Heraud, J. Beeson, D. Anderson, J. Richards, B. R. Wood, *Faraday Discuss* **2016**, *187*, 341.
[19] B. R. Wood, D. McNaughton, *J. Raman Spectrosc.* **2002**, *33*, 517.
[20] H. J. Byrne, G. D. Sockalingum, N. Stone, in *RSC Analytical Spectroscopy Series*, (Ed: D. Moss), Royal Society of Chemistry, Cambridge, **2010**, pp. 105–143.
[21] B. Zhang, B. Kuster, *Mol. Cell. Proteomics MCP* **2019**, *18*, S1.
[22] D. J. Crockford, E. Holmes, J. C. Lindon, R. S. Plumb, S. Zirah, S. J. Bruce, P. Rainville, C. L. Stumpf, J. K. Nicholson, *Anal. Chem.* **2006**, *78*, 363.
[23] L. Blanchet, A. Smolinska, *Methods Mol. Biol. Clifton NJ* **2016**, *1362*, 209.
[24] J. Kuligowski, D. Pérez-Guaita, J. Escobar, I. Lliso, M. de la Guardia, B. Lendl, M. Vento, G. Quintás, *Talanta* **2014**, *127*, 181.
[25] D. Perez-Guaita, K. Kochan, M. Martin, D. W. Andrew, P. Heraud, J. S. Richards, B. R. Wood, *Vib. Spectrosc.* **2017**, *91*, 46.
[26] P. Lasch, I. Noda, *Anal. Chem.* **2017**, *89*, 5008.
[27] P. Lasch, I. Noda, *Appl. Spectrosc.* **2019**, *73*, 359.
[28] C. Bedia, À. Sierra, R. Tauler, *Anal. Bioanal. Chem.* , DOI:10.1007/s00216-020-02595-8.
[29] J. Kuligowski, D. Pérez-Guaita, Á. Sánchez-Illana, Z. León-González, M. de la Guardia, M. Vento, E. F. Lock, G. Quintás, *Analyst* **2015**, *140*, 4521.
[26] Ferenc (2020). MIDAS 2010.(https://www.mathworks.com/matlabcentral/fileexchange/32384-midas-2010), MATLAB Central File Exchange.
[31] J. Jaumot, A. de Juan, R. Tauler, *Chemom. Intell. Lab. Syst.* **2015**, *140*, 1.
[32] A. Rygula, K. Majzner, K. M. Marzec, A. Kaczor, M. Pilarczyk, M. Baranska, *J. Raman Spectrosc.* **2013**, *44*, 1061.
[33] H. Wu, J. V. Volponi, A. E. Oliver, A. N. Parikh, B. A. Simmons, S. Singh, *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 3809.
[34] E. Szafraniec, K. Majzner, Z. Farhane, H. J. Byrne, M. Lukawska, I. Oszczapowicz, S. Chlopicki, M. Baranska, *Spectrochim. Acta. A. Mol. Biomol. Spectrosc.* **2016**, *169*, 152.
[35] M. Masetti, H. Xie, Ž. Krpetić, M. Recanatini, R. A. Alvarez-Puebla, L. Guerrini, *J. Am. Chem. Soc.* **2015**, *137*, 469.
[36] D. Pérez-Guaita, J. Kuligowski, S. Garrigues, G. Quintás, B. R. Wood, *The Analyst* **2015**, *140*, 2422.
[37] H. Stenlund, E. Johansson, J. Gottfries, J. Trygg, *Anal. Chem.* **2008**, *81*, 203.
[38] A. Barth, *Biochim. Biophys. Acta-Bioenerg.* **2007**, *1767*, 1073.
[39] D. Perez-Guaita, J. Ventura-Gayete, C. Pérez-Rambla, M. Sancho-Andreu, S. Garrigues, M. de la Guardia, *Anal. Bioanal. Chem.* **2012**, *404*, 649.
[40] B. R. Wood, *Chem. Soc. Rev.* **2016**, *45*, 1980.
[41] M. Jackson, G. Wagnieres, H. H. Mantsch, in *Encyclopedia of Spectroscopy and Spectrometry (Third Edition)*, (Eds: J. C. Lindon, G. E. Tranter, D. W. Koppenaal), Academic Press, Oxford, **2017**, pp. 479–487.
[42] D. Perez-Guaita, A. Sanchez-Illana, J. Ventura-Gayete, S. Garrigues, M. de la Guardia, *Analyst* **2014**, *139*, 170.
[43] A. Piovesan, M. C. Pelleri, F. Antonaros, P. Strippoli, M. Caracausi, L. Vitale, *BMC Res. Notes* , DOI:10.1186/s13104-019-4137-z.
[44] E. Bianconi, A. Piovesan, F. Facchin, A. Beraudi, R. Casadei, F. Frabetti, L. Vitale, M. C. Pelleri, S. Tassani, F. Piva, S. Perez-Amodio, P. Strippoli, S. Canaider, *Ann. Hum. Biol.* **2013**, *40*, 463.
[45] D. Agudelo, P. Bourassa, G. Bérubé, H.-A. Tajmir-Riahi, *Int. J. Biol. Macromol.* **2014**, *66*, 144.
[46] I. Noda, *Biomed. Spectrosc. Imaging* **2015**, *4*, 109.

[47] O. Klementieva, C. Sandt, I. Martinsson, M. Kansiz, G. K. Gouras, F. Borondics, *Adv. Sci.* **2020**, *7*, 1903004.
[48] Q. Zhang, P. Zhang, H. Gou, C. Mou, W. E. Huang, M. Yang, J. Xu, B. Ma, *The Analyst* **2015**, *140*, 6163.
[49] J. Zhou, Z. Wang, S. Sun, M. Liu, H. Zhang, *Biotechnol. Appl. Biochem.* **2001**, *33*, 127.

Figure Captions

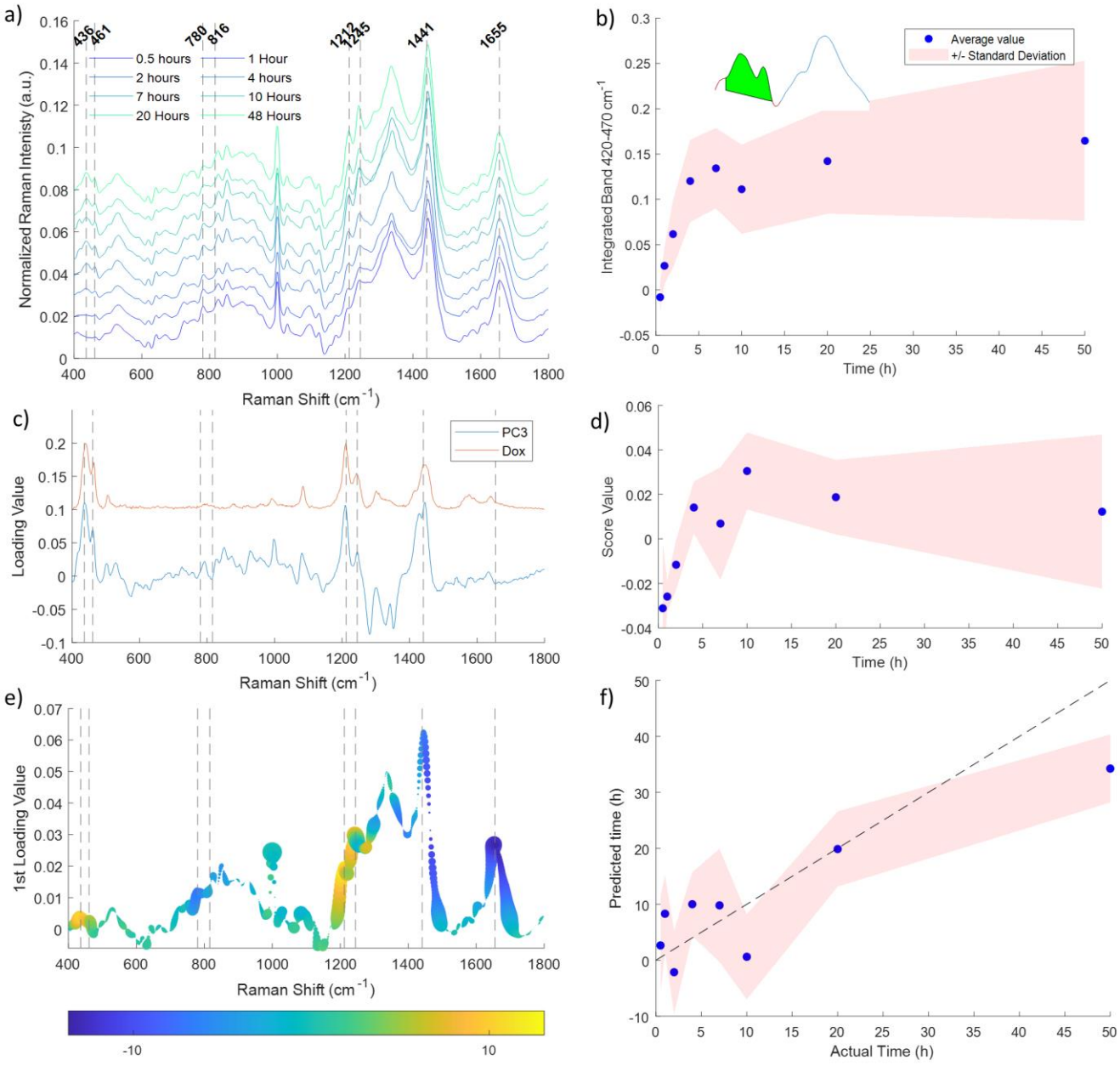


Figure 1. *Raman study of DOX uptake. (a) Average Raman spectra of cells inoculated at different timepoints. (b) Temporal evolution of the integrated the 420-470 cm^{-1} band. Shading indicates the standard deviation. (c) Spectral Loading of PC3 and Raman spectra of a DOX aqueous solution standard. (d) PC3 Average Score values over time. (e)*

Scatter plot indicating the average Raman spectra. The colour scale indicates the component loading of the oPLSR and the size of the points is proportional to the selectivity ratio. (f) Actual time versus the time predicted by the cross validation of the oPLSR. The dashed line depicts the idealised 1:1 correlation.

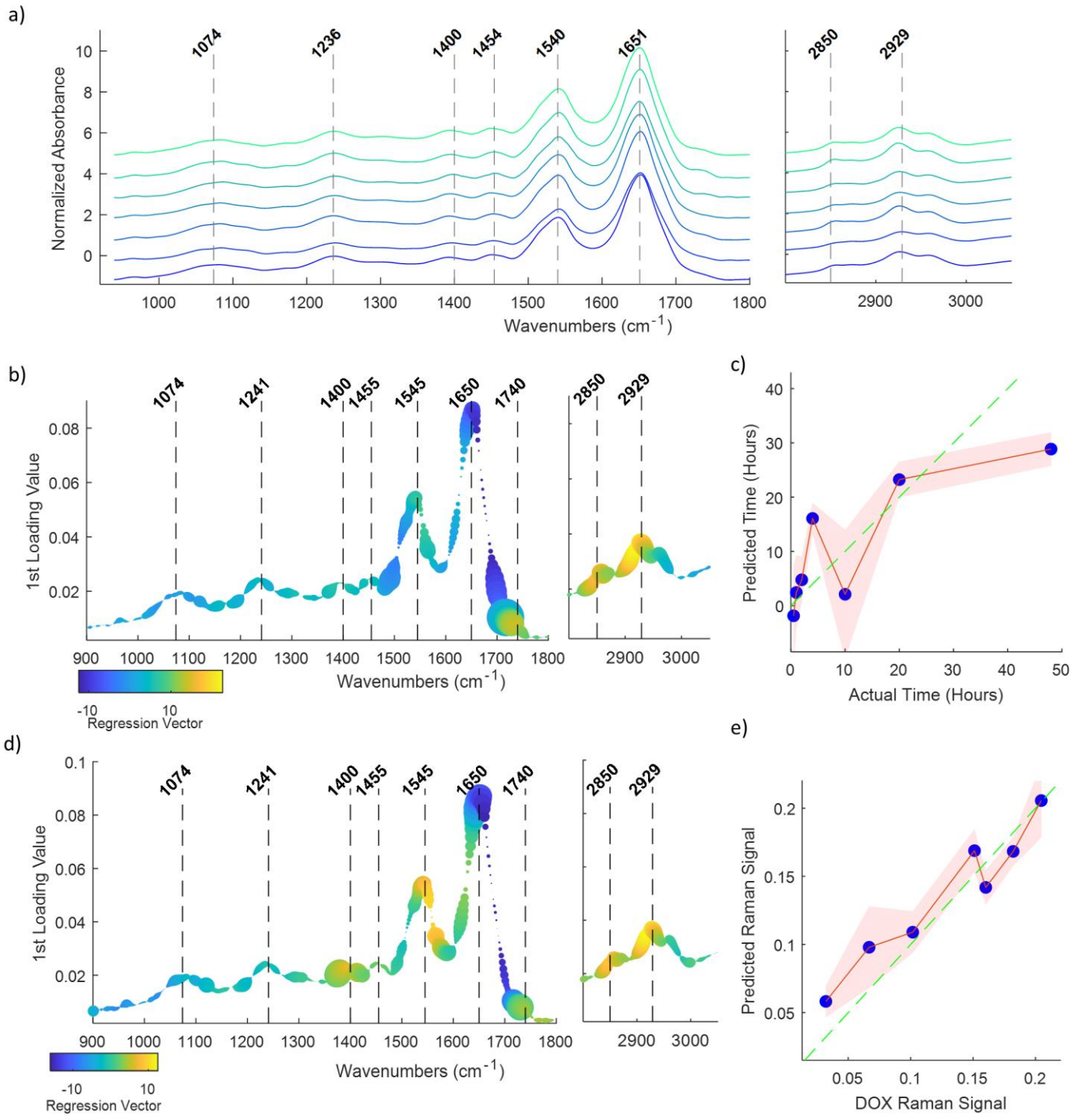


Figure 2. *Infrared study of DOX uptake by cells. (a) Average infrared spectra of cells inoculated at different timepoints. (b) Analysis of variables of the oPLSR which predicted the incubation time. The scatter plot indicates the average infrared spectra, the colour scale indicates the component loading of the oPLSR and the size of the points is proportional to the selectivity ratio. (c) Actual time versus the time predicted by the cross validation of the oPLSR model which predicted the incubation time. (d) Analysis of variables of the oPLSR model which predicted the 430-482 cm^{-1} Raman band. The scatter plot indicates the average infrared spectra, the colour scale indicates the component loading of the oPLSR and the size of the points is proportional to the selectivity ratio. The dashed line depicts the idealised 1:1*

correlation.(e) Actual integrated 430-482 cm^{-1} Raman band versus the predicted by the cross validation of the oPLSR model.

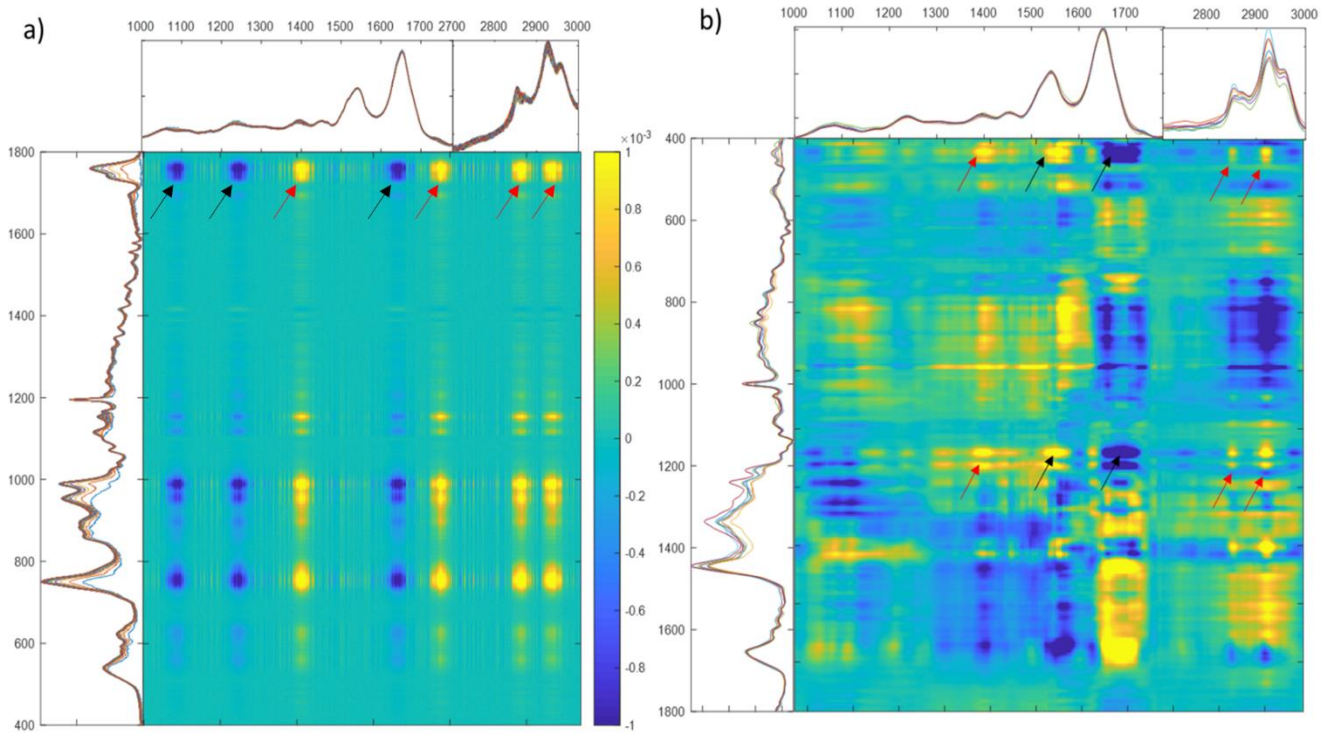


Figure 3. *Synchronous coefficients obtained from the 2D correlational joint analysis of Raman and IR data. (a) Simulated Dataset. (b) Experimental Dataset.*

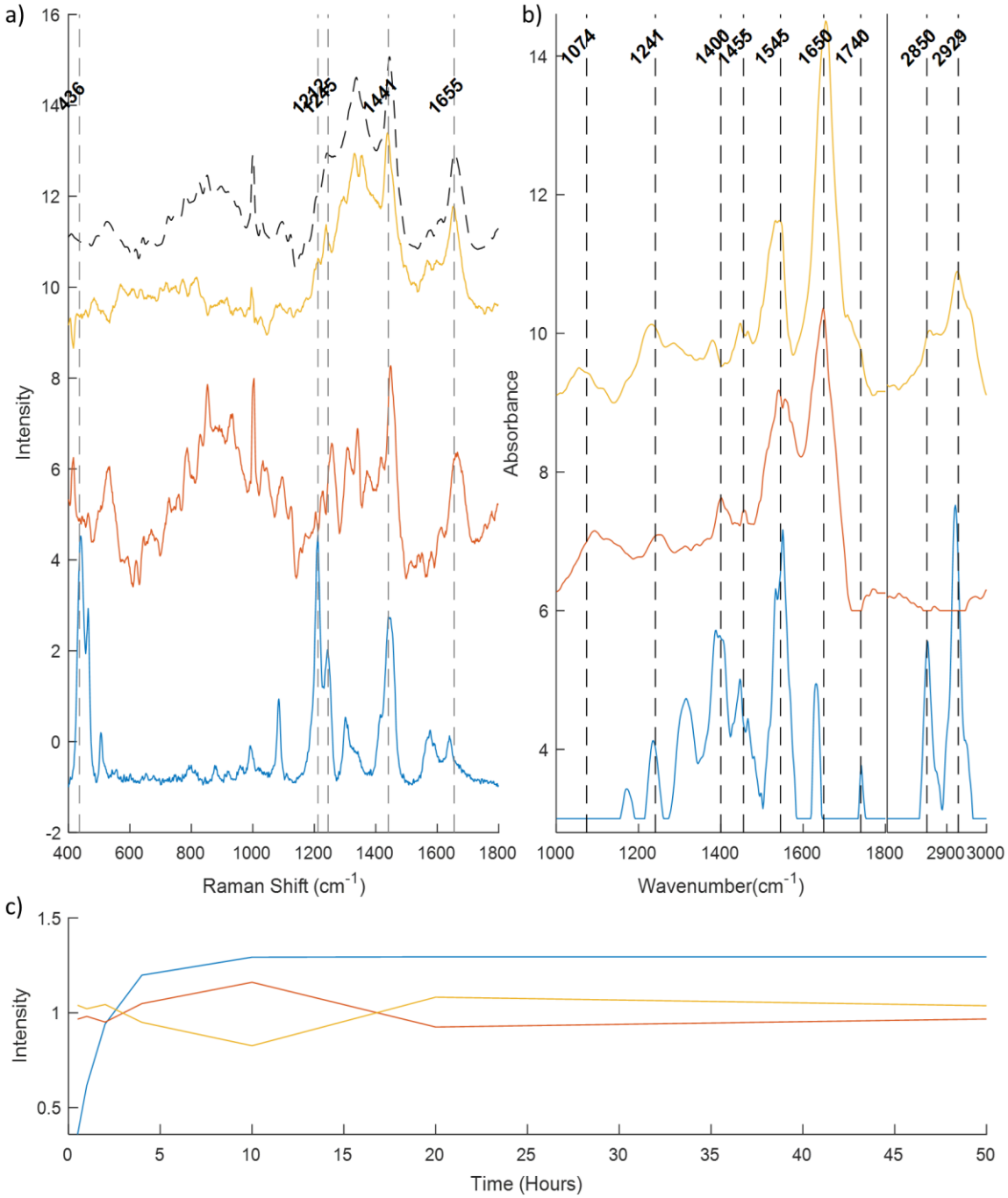


Figure 4. *MCR-ALS results for the joint analysis of IR and Raman datasets. (a) Pure Raman spectra of the three components. Black dotted line represents the Raman spectra of a cell (Shortest incubation time). (b) Pure Infrared spectra of the three components considered. (c)Evolution of the concentration profiles over time of the three components considered.*

Supplementary Material

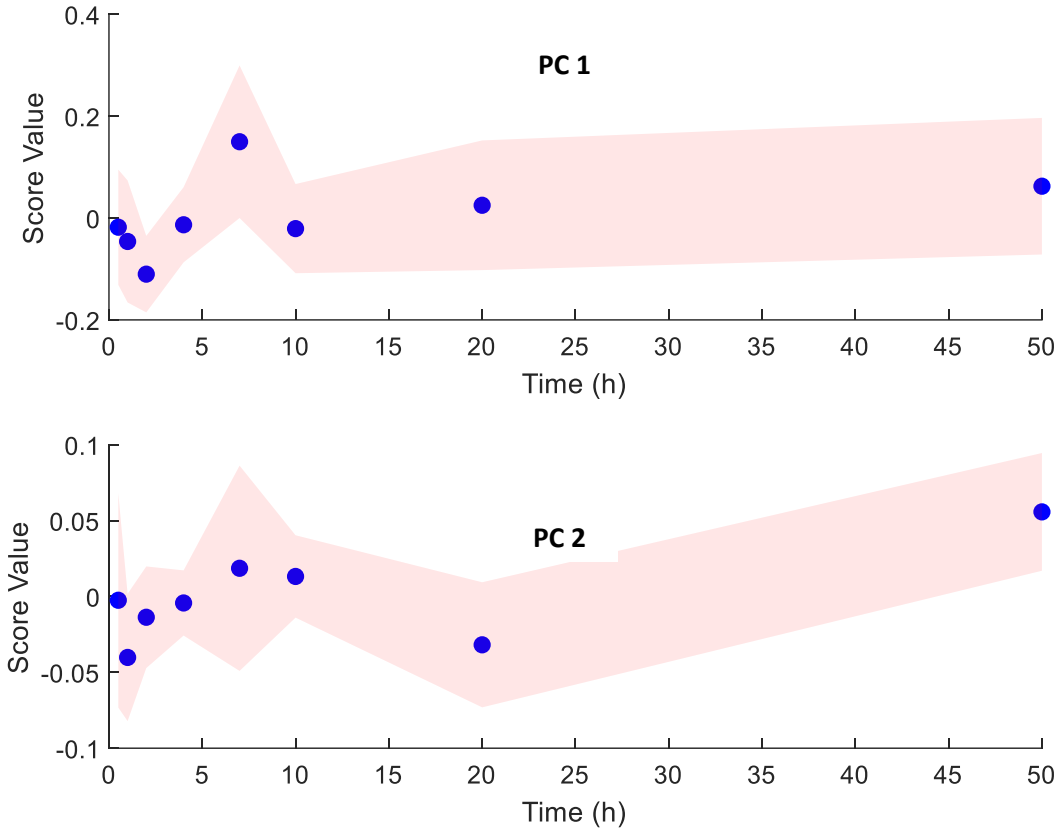
MULTIMODAL VIBRATIONAL IN VITRO STUDIES OF DRUG UPTAKE: IS THE WHOLE GREATER THAN THE SUM OF THEIR PARTS?

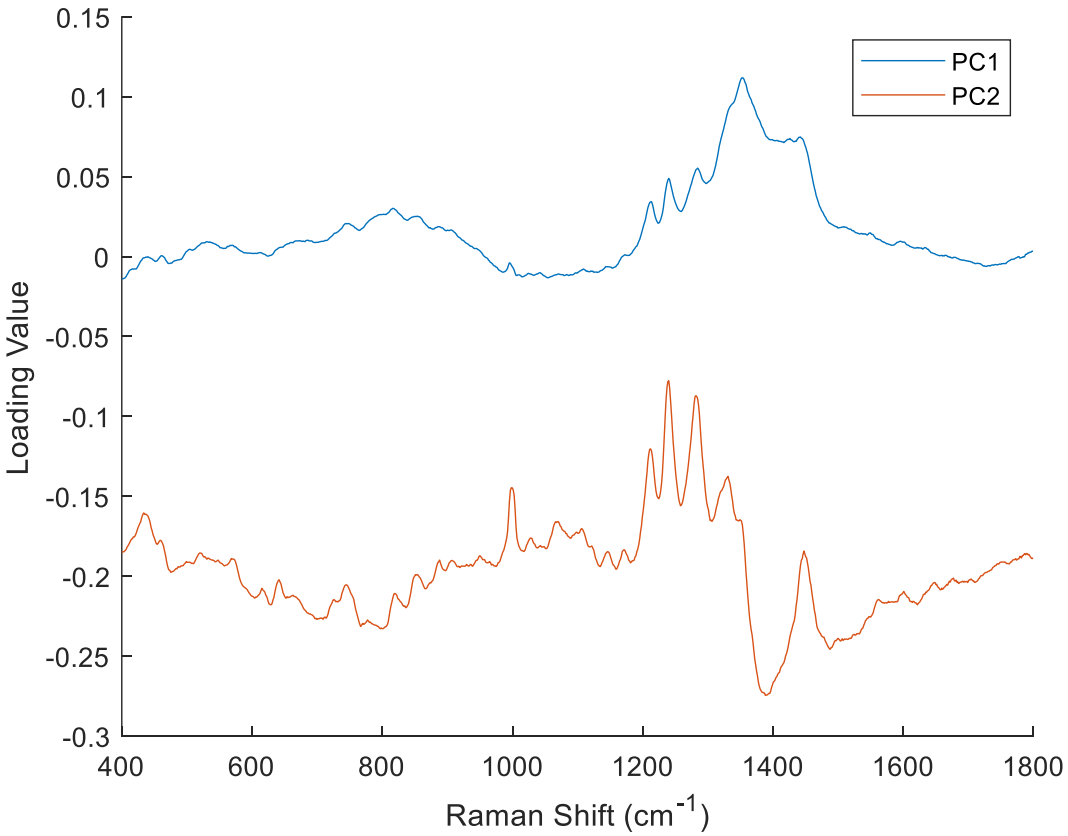
David Perez-Guaita*[1], Karolina Chrabaszcz[2], Kamilla Malek[2], Hugh J. Byrne[1]

[1]*FOCAS Research Institute, Technological University Dublin, City Campus, Dublin 8, Ireland*

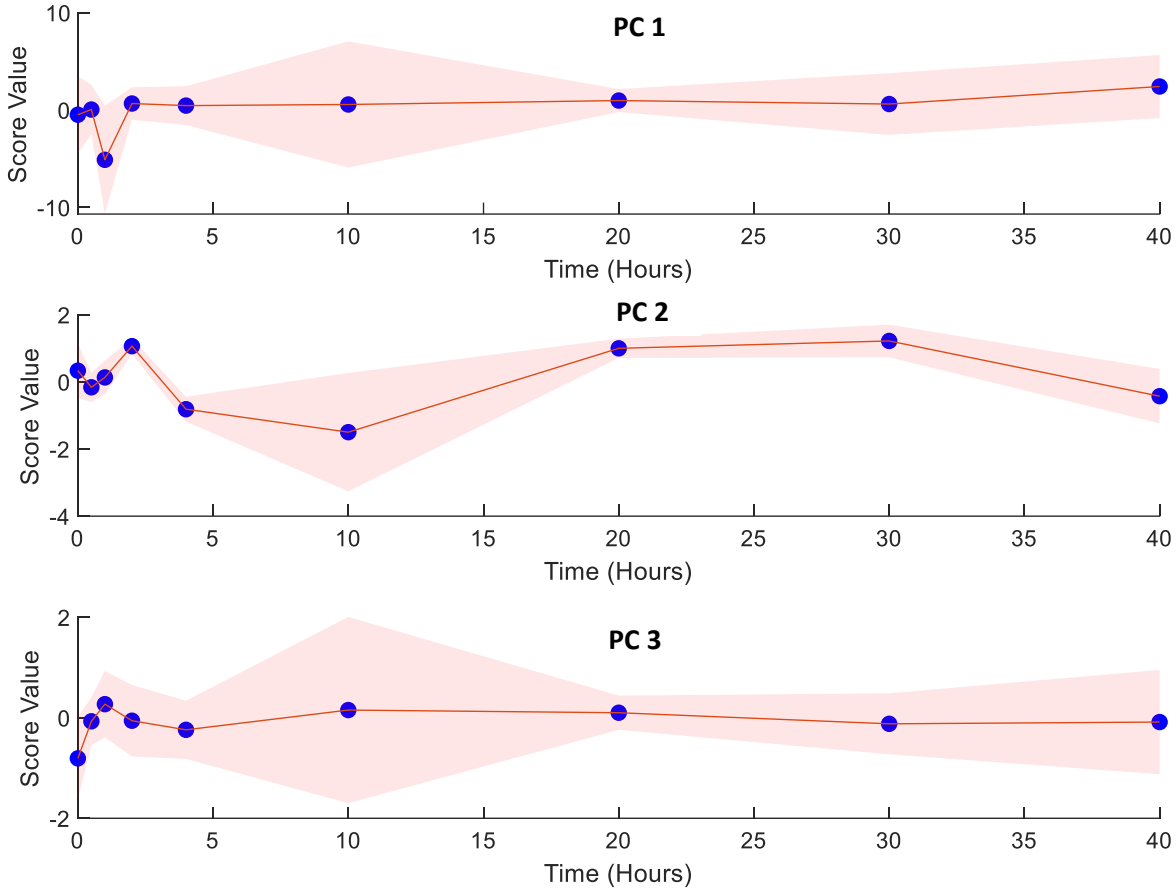
[2] *Faculty of Chemistry, Jagiellonian University, Krakow, Poland*

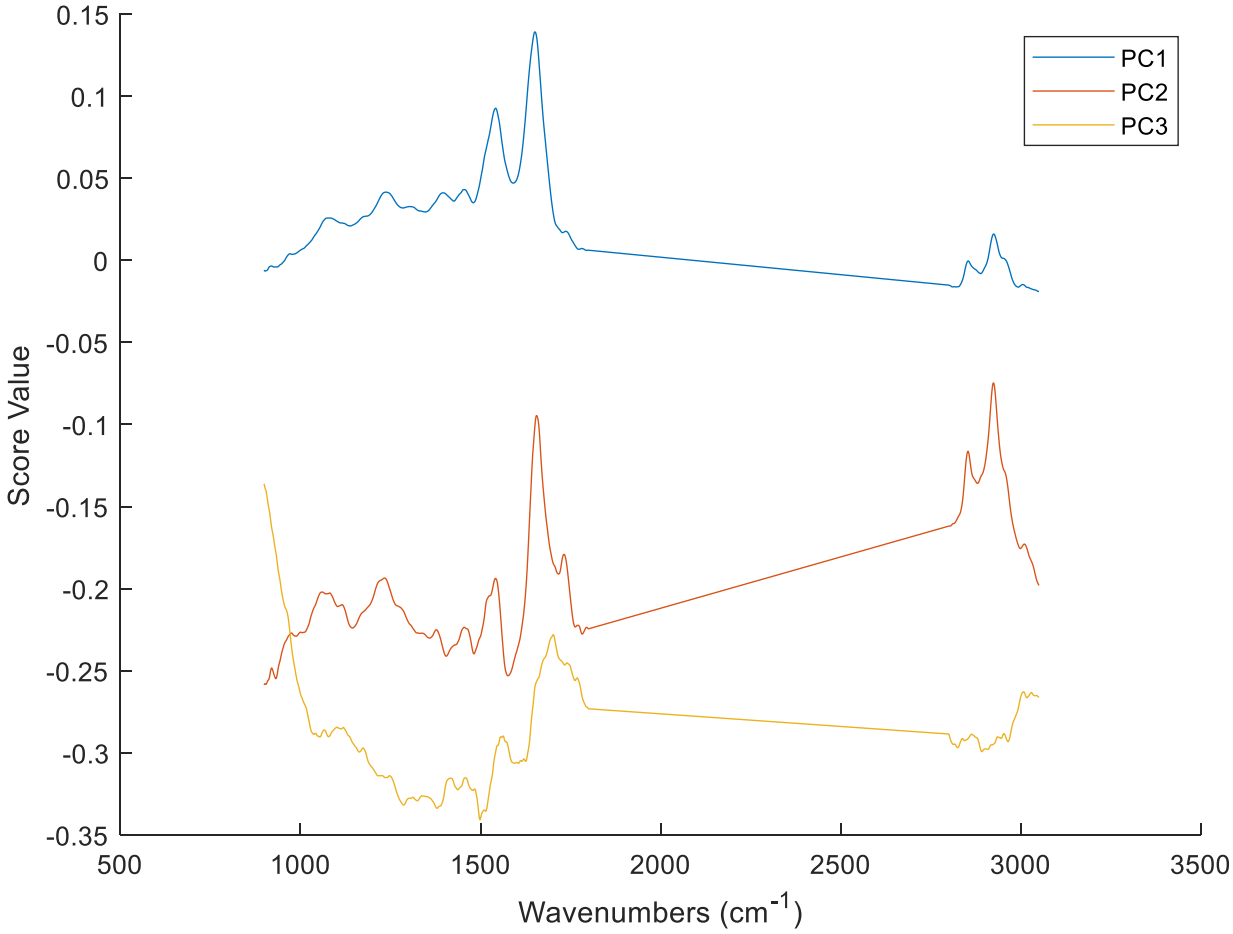
SM1: Score and loading values corresponding to Principal Components #1 and #2 obtained from the PCA of the Raman Dataset.



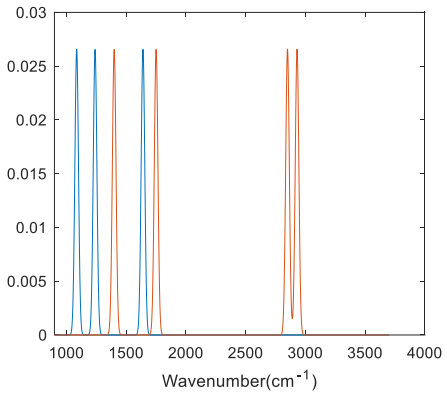


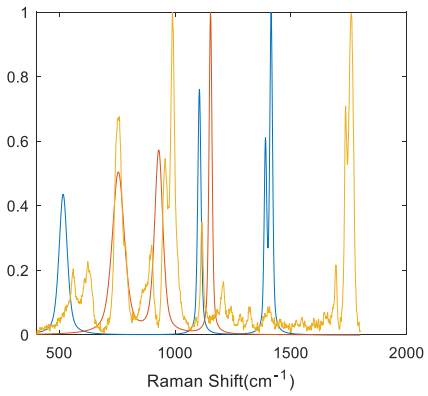
SM2: Score and loading values corresponding to Principal Components #1, #2 and #3 obtained from the PCA of the IR Dataset.

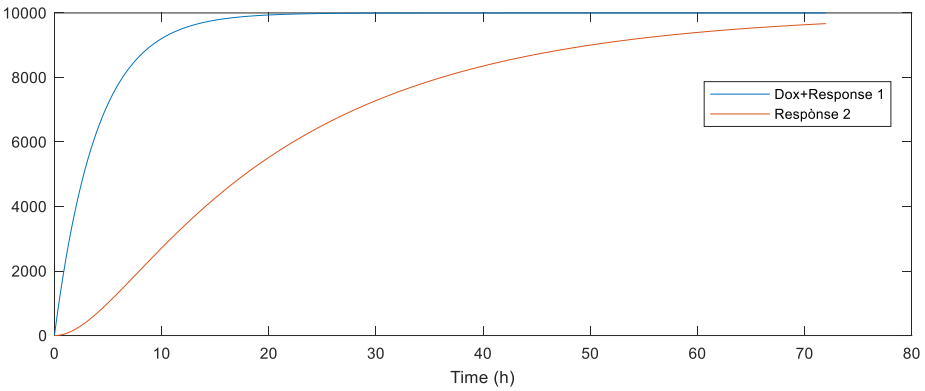




SM3: Simulation of the IR and Raman spectrum. Artificial construct added to the spectra for IR (a) and Raman (b). (c) Intensity of the two effects added to the spectra. For more information see [9,37].







SM.4 Asynchronous analysis for simulated and real data.